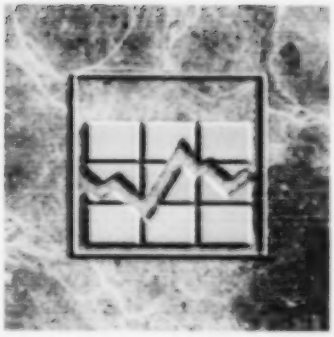
N° 75F002MIF au catalogue — N° 003
ISSN : 1707-2867
ISBN : 978-0-662-08216-3

Document de recherche

Série de documents de recherche - Revenu

Renouvellement de l'échantillon de l'enquête sur la dynamique du travail et du revenu



par Sylvie Michaud, Lecily Hunter and Georges Lemaître

Division de la statistique du revenu
Immeuble Jean-Talon, Ottawa, K1A 0T6

Téléphone : 1-613-951-7355

Statistique Canada Statistics Canada

Canada

RENOUVELLEMENT DE L'ÉCHANTILLON DE L'ENQUÊTE SUR LA DYNAMIQUE DU TRAVAIL ET DU REVENU

Mars 1992

Sylvie Michaud, Division des méthodes d'enquêtes sociales

Lecily Hunter, Division des méthodes d'enquêtes sociales

Georges Lemaître, Division des enquêtes des ménages

Les Documents de recherche de l'Enquête sur la dynamique du travail et du revenu sont conçus en vue de communiquer les résultats des études ainsi que les décisions importantes prises à l'égard de cette enquête. Ils sont offerts gratuitement, en français et en anglais. Pour obtenir une description sommaire des documents disponibles ou un exemplaire de ces documents, communiquer avec Philip Giles, EDTR, Édifice Jean-Talon, 11ième étage, section D8, Statistique Canada, Ottawa (Ontario), Canada, K1A 0T6.
Téléphone: (613) 951-2891; Télécopieur: (613) 951-3253.

SOMMAIRE

Parce que l'EDTR est une enquête longitudinale, il est souhaitable que les répondants fassent partie de l'échantillon le plus longtemps possible. Toutefois, l'érosion de l'échantillon et les changements qui surviennent au sein de la population rendent l'échantillon de moins en moins représentatif de la population au fur et à mesure que le temps passe. Pour équilibrer ces facteurs, l'échantillon d'une étude longitudinale est habituellement formé de panels représentatifs de la population cible. Cet article examine les options possibles quant au nombre de panels dans l'échantillon et à la durée de la période que chaque panel doit participer à l'enquête, étant donné un échantillon de fixe. Il comprend en outre une explication des raisons pour la décision du choix final.

1. INTRODUCTION

L'enquête sur la dynamique du travail et du revenu (EDTR) est la première enquête longitudinale sociale de Statistique Canada. La collecte des données doit commencer en janvier 1993. La taille de l'échantillon a été fixée à 40 000 ménages. Toutefois, le plan d'échantillonnage doit permettre d'établir un compromis entre de nombreux facteurs :

a) Exigences longitudinales versus exigences transversales

Pour l'analyse longitudinale, la durée des périodes pendant laquelle les unités font partie de l'échantillon doivent être aussi longues que possibles. Toutefois, on s'attend que l'enquête produise certaines estimations transversales, dont la fiabilité diminuerait chaque année en raison des variations de la population et de la perte d'effectifs de l'échantillon. Il est donc nécessaire de procéder à un certain remplacement de l'échantillon. Pour effectuer une analyse historique, il faut que la durée de la période en cause soit «raisonnablement longue» étant donné que les événements qui nous intéressent ne se produiront pas opportunément au beau milieu de la période de collecte. L'allongement de la durée du panel permet d'améliorer la qualité des résultats au niveau de la modélisation et elle accroît les chances de recueillir des renseignements de base additionnels. Les questions de fond suivantes ne peuvent être convenablement étudiées si la période d'observation est trop brève :

- périodes de chômage révolues et somme des périodes de chômage au fil des ans;
- périodes de pauvreté révolues;
- événements qui ont précédé et suivi la séparation d'un couple;
- facteurs déterminants au plan de la productivité et de la compétence du travailleur;
- stabilité au niveau de l'emploi;

- ! conséquences de la maladie sur l'emploi et le revenu;
- ! transitions dans le cycle de vie;
- ! fluctuations de la richesse.

En tentant de balancer la taille de l'échantillon et la durée d'un panel, ce deuxième point est beaucoup plus important du point de vue de l'analyse longitudinale. Nous sommes d'avis que cette durée devrait être d'au moins cinq ans.

b) Principales utilisations des données

Comme nous l'avons mentionné, nous devons procéder à la fois à des analyses transversales et longitudinales. En outre, il faut également mener certaines analyses à l'échelle provinciale et à d'autres niveaux infranationaux.

c) Taux de réponse prévus

Plus les taux de non-réponse prévus sont élevés, plus grande doit être la taille de l'échantillon afin de produire des estimations de fiabilité égale. L'incidence de la non-réponse sur le biais des estimations constitue un autre facteur.

d) Fardeau de réponse

Celui-ci entre en ligne de compte non seulement en raison de son incidence sur une interview en particulier mais également sur celles qui suivront. On constatera certainement une perte d'effectifs de l'échantillon chaque fois qu'on mène des interviews étant donné que des répondants décideront alors de ne plus participer à l'enquête.

2. HYPOTHÈSES DE TRAVAIL

a) La taille globale de l'échantillon s'établira à 40 000 ménages. Toutefois, à l'origine on prévoyait qu'elle atteindrait 50 000 ménages. On étudie en l'occurrence la possibilité d'élargir la taille de l'échantillon à un moment donné.

b) L'échantillon initial de l'EDTR sera constitué d'un sous-échantillon de l'Enquête sur la population active (EPA). À moins de problèmes de réponse qui découleraient de l'utilisation de l'échantillon de l'EPA, les prochains panels seront sélectionnés de la même manière.

c) L'EDTR comportera une interview préliminaire et des interviews annuelles périodiques. La première interview sera menée comme enquête supplémentaire de l'EPA. (Veuillez noter que, pour ce faire, il faut que l'échantillon soit composé de ménages de l'EPA, tel qu'indiqué en (b) ci-dessus. Si on choisit un échantillon composé d'autres ménages que ceux de l'EPA, l'interview préliminaire devra être menée séparément. Il faudra alors remanier son contenu de façon à pouvoir demander les données recueillies dans le contexte de l'EPA. Les conséquences financières sont évidentes.) Par conséquent, les répondants seront interviewés (par téléphone si possible) deux fois par année : en janvier, pour recueillir les renseignements sur le marché du travail (selon le modèle de l'Enquête sur l'activité (EA) et en mai, pour recueillir des données sur le revenu (en utilisant une version modifiée de l'Enquête sur les finances des consommateurs (EFC)). (On peut obtenir des détails additionnels sur la décision concernant les dates d'interview du Document de recherche de l'EDTR 92-02, «Enquête sur la dynamique du travail et du revenu: Dates d'interview possibles».)

d) Il se peut qu'on mène chaque mai une enquête supplémentaire (à l'EDTR) sur un sous-échantillon de 10 000 ménages.

e) Le taux de non-réponse et le taux de perte d'effectifs sont semblables à ceux obtenus par le Bureau of the Census des États-Unis pour SIPP (Survey of Income and Program Participation, semblable à l'EDTR). Les taux réels de l'enquête américaine sont présentés à l'annexe A. Ils révèlent que le taux de perte d'effectifs est plus élevé dans le cas des premiers panels, le taux de perte d'effectifs supplémentaire s'élevant à 1.8 % chaque fois qu'une interview est menée. Toutefois, le taux de réponse de l'EA s'établit aux environs de 92 % et celui de l'EFC est tout juste inférieur à 80 %. Par conséquent, les taux de non-réponse de l'EDTR seront peut-être supérieurs à ceux supposés en l'occurrence.

f) En moyenne, les ménages comprennent deux personnes éligibles à répondre à l'enquête.

3. OPTIONS RELATIVES AU RENOUVELLEMENT DE L'ÉCHANTILLON

Nous avons étudié plusieurs plans de renouvellement de l'échantillon. Certains n'ont toutefois pas été examinés en profondeur. En vertu d'une des options, la totalité des 40 000 ménages ferait partie de l'échantillon pendant cinq ans et serait ensuite complètement remplacée. Cette option comporterait d'importantes contraintes opérationnelles au moment du renouvellement de l'échantillon. En outre, ce plan de renouvellement serait, du point de vue analytique, vulnérable aux modifications apportées aux programmes gouvernementaux, les récessions économiques, etc. Pour ce qui est de la méthodologie, la détérioration de l'échantillon ferait que les données transversales seraient bien meilleures certaines années que d'autres. Par conséquent, nous en sommes venus à la conclusion qu'il est hautement souhaitable d'adopter un plan de renouvellement sur une partie seulement de l'échantillon.

Nous n'avons examiné aucune option qui nous obligerait à modifier considérablement chaque année la taille de l'échantillon. Par exemple, dans le cas d'une approche préconisant deux panels qui feraient tous deux partie de l'échantillon pendant cinq ans, la taille de l'échantillon transversal s'établirait alors à 40 000 et, l'année suivante, à 60 000 et ainsi de suite. Bien que cette façon de procéder ne pose aucun problème majeur du point de vue de l'analyse longitudinale, la fiabilité des données transversales varierait chaque année.

Finalement, nous avons étudié attentivement trois plans de renouvellement de l'échantillon.

Dans les plans de renouvellement soumis, P désigne l'interview préliminaire tandis que le chiffre indique le nombre d'années pendant lesquelles le panel a fait partie de l'échantillon jusqu'à présent (par exemple, 3 signifie qu'il s'agit de la troisième d'année d'interview).

Option A : 40 000 ménages avec renouvellement annuel de 1/5 de l'échantillon (8 000 ménages)

L'échantillon transversal de 1993 serait introduit graduellement, en commençant par deux groupes de renouvellement de l'EPA. L'échantillon transversal intégral comprendrait 40 000 ménages. Un panel longitudinal intégral (par exemple, celui de 1994-1998) compterait 8 000 ménages).

Panel	1993	1994	1995	1996	1997	1998	1999	2000	2001...
A	P	1	2	3	4				
B	P	1	2	3	4	5			
C		P	1	2	3	4	5		
D			P	1	2	3	4	5	
E				P	1	2	3	4	5
F					P	1	2	3	4
G						P	1	2	3
H							P	1	2

Avantages :

- Ce plan de sondage offrirait aux utilisateurs des données longitudinales pour 5 ans.
- Le plan de rotation serait symétrique, en ce sens que la taille de l'échantillon et, par conséquent, la tâche de l'intervieweur seraient constantes d'année en année (1/6 d'interviews personnelles et 5/6 d'interviews téléphoniques).
- Pour les estimations transversales, une partie du biais attribuable à la perte d'effectifs pourrait être compensée par le renouvellement constant de l'échantillon.
- Si on augmentait à 50 000 ménages la taille de l'échantillon, il serait facile de modifier le plan de sondage (sans hausse majeure des frais) puisqu'un groupe de renouvellement de l'EPA comprend environ 10,500 ménages.

Désavantages :

- Après cinq ans, le taux de non-réponse global pourrait se situer aux environs de 33 %, en supposant que la perte d'effectifs de l'échantillon soit

semblable à celle de SIPP (se stabilisant à environ 1.8 % après les trois premières interviews). Il resterait alors 10,700 personnes dans l'échantillon longitudinal. Ce nombre serait peut-être suffisant pour procéder à certaines analyses au niveau national mais il ne permettrait pas de se livrer à des analyses à l'échelle provinciale.

- Même si la perte d'effectifs de l'échantillon était faible, l'échantillon restreint de chaque panel entraînerait une réduction de l'ampleur des analyses longitudinales.
- En supposant que des enquêtes supplémentaires soient menées et que celles-ci contribuent à la perte d'effectifs, le taux de non-réponse pourrait être aussi haut que 40 %, à la fin de cinq années.

Option B : <u>40 000 ménages avec rotation bisannuelle de 20 000 ménages</u>

Panel -	1993	1994	1995	1996	1997	1998	1999	2000	2001...
A	P	1	2	3	4				
B			P	1	2	3	4		
C					P	1	2	3	4
D							P	1	2

Ce plan d'échantillonnage est celui qui comporterait le plus grand nombre de données longitudinales pour une période de quatre ans. À tous les deux ans, la moitié de l'échantillon serait renouvelée.

Avantages :

- Grâce à ce plan de renouvellement, chaque panel comprendrait plus de données longitudinales. (Veuillez noter que 20 000 ménages feraient partie

de l'échantillon pendant 4 ans comparativement à 16 000 pour ce qui est de l'option A.)

- Les répondants seraient interviewés pendant 5 ans et non pas 6 ans, ce qui diminuerait le fardeau de ces derniers.
- Étant donné qu'un panel comprend 20 000 ménages, il serait possible de sélectionner deux groupes de rotation de l'EPA, ce qui signifie que cette approche serait relativement peu coûteuse.
- S'il fallait modifier ultérieurement le contenu de l'EDTR, il serait possible de modifier un panel à la fois. Chaque demi-échantillon serait suffisamment grand pour permettre de procéder à certaines analyses.

Désavantages :

- Des données longitudinales complètes ne seraient disponibles que pour une période de quatre ans seulement.
- Tous les deux ans, il faudrait mener un bon nombre d'interviews personnelles. La charge de travail des intervieweurs ne serait donc pas équilibrée.
- S'il fallait obtenir des données longitudinales pour une période de cinq ans, ce plan de renouvellement ne serait pas symétrique et la taille de l'échantillon transversal varierait considérablement d'année en année.
- Si la taille de l'échantillon augmentait pour atteindre 50 000 ménages, la sélection de l'échantillon serait probablement plus coûteuse. En fait, il semble probable que, dans les régions rurales, il serait difficile d'augmenter le ratio d'échantillonnage dans les unités primaires d'échantillonnage. Il faudrait alors lister des grappes additionnelles. (Une autre option consisterait à sélectionner un sous-échantillon de trois groupes de rotation, ce qui serait toutefois trop onéreux.)

Option C : <u>40 000 ménages avec renouvellement de la moitié de l'échantillon (20 000 ménages) tous les 2 ou 3 ans, selon les taux de réponse obtenus</u>

Il s'agit d'une version modifiée de l'option B. Ce plan comporte l'avantage de déterminer en 1997 la durée d'utilisation d'un panel, en se basant sur les taux de réponse. Dans l'optique de ce plan, l'année 1997 constitue donc un point tournant : si les taux de réponse semblent alors se détériorer trop rapidement pour que le panel soit utilisé pendant plus de quatre ans, le renouvellement de l'échantillon s'effectuerait tous les deux ans. Toutefois, si les taux de réponse se stabilisent à un niveau acceptable, le renouvellement serait effectué tous les trois ans, ce qui permettrait d'obtenir des données longitudinales pour une période de six ans.

Si les taux de réponse révèlent que la perte d'effectifs de l'échantillon est trop élevée pour interviewer les mêmes répondants pendant six ans, un nouvel échantillon serait adopté en 1998 et les prochains panels seraient maintenus pendant quatre ans. L'échantillon B serait toutefois suivi pendant cinq ans. (Ceci est important pour pouvoir mener l'enquête supplémentaire.)

Renouvellement de l'échantillon aux deux ans

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003...
Panel											
A	P	1	2	3	4	5					
B				P	1	2	3	4			
C						P	1	2	3	4	
D								P	1	2	3
E										P	1

Renouvellement de l'échantillon aux trois ans

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002...
Panel										
A	P	1	2	3	4	5	6			
B				P	1	2	3	4	5	6
C							P	1	2	3
D										P

Avantages :

- Ce plan comporte la possibilité de déterminer la durée d'utilisation du panel en fonction des taux de réponse observés à l'origine.
- S'il était possible d'obtenir six ans de données longitudinales, la charge de l'intervieweur serait réduite étant donné l'introduction d'un nouveau panel aux trois ans plutôt qu'aux deux ans (qui nécessite des interviews personnelles).
- Pour ce qui est de l'analyse longitudinale, il s'agit du meilleur plan car la taille du panel de même que la durée de période de collecte seraient supérieures à celles des options A ou B.
- S'il fallait modifier le contenu, l'introduction des panels pourrait se faire séparément, ce qui éviterait de troubler la définition des fichiers des répondants. Parallèlement, chaque panel serait suffisamment grand pour permettre de procéder à certaines analyses.

Désavantages :

- La collecte de données annuelles sur le travail et le revenu pendant une période de six ans représenterait au total sept ans d'interviews, ce qui serait un lourd fardeau pour les répondants.
- Il se pourrait que le taux de perte d'effectifs soit élevé.
- Comme pour l'option B, si la taille de l'échantillon augmentait pour se fixer à 50 000 ménages, on constaterait vraisemblablement une hausse des coûts d'établissement des listes.

4. CONCLUSIONS

En dépit du fait que l'EDTR doit permettre d'obtenir certaines données transversales, il s'agit d'abord et avant tout d'une enquête longitudinale. Nous avons jugé que l'option C, qui prévoit un renouvellement de l'échantillon tous les trois ans, est la meilleure façon de soutenir l'analyse longitudinale. En outre, il sera plus facile d'apporter des modifications en même temps qu'on renouvelle l'échantillon dans le cas où les HYPOTHESES de planification seraient erronées.

Annexe A

Taux de réponse et perte d'effectifs de l'échantillon pour SIPP

(basés sur le panel de 1984,

personnes admissibles)

Répondants aux 8 entrevues		72.2 %
Perte d'effectifs totale		19.6 %
Perte d'effectifs par vague		
- Vague 2	4.3 %	
- Vague 3	3.7 %	
- Vague 4	3.0 %	
- Vague 5	2.8 %	
- Vague 6	2.2 %	
- Vague 7	1.9 %	
- Vague 8	1.8 %	
Autre non-réponse		8.1 %